IOSR Journal of Computer Engineering (IOSRJCE)
ISSN: 2278-0661 Volume 3, Issue 4 (July-Aug. 2012), PP 07-10
www.iosrjournals.org

Substitution Error Analysis for Improving the Word Accuracy in Telugu Language Automatic Speech Recognition System

M. Nagamani, P. N. Girija
University of Hyderabad, Hyderabad.

Abstract: *Use of natural languages for the computer communication is one of the current research topics. The speech recognition plays a central role in communicating the computer by means of speech. Speech Recognition is the process of converting analog signal into the symbolic gesture form known as text. An Automatic Speech Recognition (ASR) is the process of converting input speech signal given to the system, into text. This input signal is any human spoken word. Though the last Four decades, research is going on bringing the system perception near to the human being, in recognition word accuracy. Many domains play a role in degrading the of system performance in which language and its pronunciation variants caused by different reasons like accent, gender, dialects of the speech are in general factors. In specific, system environment, articulatory phonetics, acoustic phonetics, (acoustic system), lexical model (pronunciation dictionary) and language model including the mood of the speaker. Acoustic modeling can be done in most of the cases by using signal processing, where as lexicon model require a sufficient human intervention as, it is based on the language and human perception. Once sufficient primary data is built then automatic processing can be done using different modeling techniques to derive more data for proper training of the speech recognition system. Here the linguistics are play a major role in building the robust system.*

In lexical model design language plays a major role. Each language has its own rhythm in speech and language aspects. Based on language rhythm worldly languages are classified as stress timed and syllable timed rhythm. Most of ASR systems use the lexical model that is built for stressed timed languages. All Indian languages are syllable timed rhythmic languages one such is Telugu Language. In this paper analysis of the decoding results of ASR system using two different lexical model environments. One is CMU lexicon which is based on stress timed language as the tool is used American accent English phonemes and another UOH lexicon which is handcraft lexicon for Telugu language which is also a syllable timed language.. Further studied are the gender and accents (pronunciation variant factors) effecting the Substitutional errors in ASR system. The confusion matrix for vowel and consonants alone analyzed for both cases and also for isolated word recognition where the confusion matrix gives the most common phonemes substituted. In all the cases the UOH lexicon based ASR system gives the improvement of word accuracy around 20 to 30%.

Speech is a process used to communicate from a speaker to listener. Pronunciation relates to speech, and humans have an intuitive feel for pronunciation. For instance, people chuckle when words are mispronounced and notice when foreign accent colors a speaker's pronunciations.[1]. If the words were always pronounced in the same way, ASR would be relatively easy. However, for various reasons words are almost always pronounced differently and varied from one speaker to another and from once situation to another. The variability is due to co-articulation, reasonal accents, speaking rate, speaking style etc.

I. Language need in ASR study:

There are 6912 languages are there the purpose of communication between human being around the world. The need of the language is to computerization of many human need domains, Ubiquitous information access, phone-based information access, mobile devices which demand speech as modality to interact, globalization in cross-cultural human-human interaction, multilingual communities, Humanitarian needs like disaster, healthcare, military applications to communicate with local people and main focus on Human machine interaction where people expect speech-driven applications in their mother tongue will specifically demand the speech recognition research to work in regional language . such one application here we like to develop a speech recognition system working in Telugu Language.

II. Telugu Language and ASR

According to the 1997 senses total 69,600,000 population in India who are speaking Telugu. Population tatal all countries around 69,758, 890 are the Telugu language speaking people. The language coverage regions are Andhra Pradesh and neighboring states. Also in Bahrain, Canada, Fiji, Malaysia(Peninsular), Mauritius, Singapore, South Africa, united Arab Emirates and United States. This Telugu language also called with alternate names as Andhra, Gentoo, Tailangi, Telangire, Telegu, Telgi, Tengu,

Terangi, Tolangan. The dialects of this language are Berad, Dasari, Dommara, Golari, Kamathi, Komtao, Konda-Reddi, Salewari, Telangana, Telugu, Vadaga, Srikakula, Vishakhapatnam, East Godaveri, Rayalseema, Nellore, Guntur, Vadari, Yanadi. This also classified as Dravidian, South-central and Telugu. There are 5,000,000 second language speakers. Fully developed Bible from 1854-2002. That is importance of Telugu language.

Telugu is one of the oldest language in the world. The old form of Telugu dates back to 1000BC. Approximately 74 million people speak Telugu as their native language. It is the third most widely spoken language in India. It is a language primary spoken in south India. It belongs to Dravidian group of language. Including non native speakers there are nearly 90 million people speaking Telugu in Andhra Pradesh alone. Also Telugu speakers available in USA, UK, Malaysia, Fiji Islands, Mauritius and South Africa. Nature of Telugu language is Syllabic as similar to the languages of India. Each symbol in Telugu script represent a complete syllable. There is a very little scope for confusion and speaking problems contest in Telugu unthinkable, since everyone will score 100%. In that sense it is a WYSIWYG script. This form of script is considered to be most scientific by the linguists. This syllabic script has been achieved by the use of a set of basic symbols, a set of modifier symbols and rules for modification. Officially there are eighteen Vowels, thirty six consonants and three dual symbols. Only thirteen vowels, thirty five consonants are in common usage. Telugu script has the capacity to represent almost the entire phonetic spectrum of all Indian (and most world) languages. Telugu stands as one of the best script in the world while maintaining an extensive sound base. The Telugu basic symbol pronunciation table along with the Romanization of Telugu sounds i.e RIT form of Telugu pronunciation using Telugu Lipi standards is shown in Table 1. Telugu Vuccharana pattika(Pronunciation Table)

Telugu aksharaalu (alphabets), Romanization & their Pronunciation

అ	a	son	క	ka	cart	త	ta	French t
ఆ	aa	master	ఖ	kha	blockhead	థ	tha	thumb
ఇ	i	if	గ	ga	goat	ద	da	then
ఈ	ia	feel	ఘ	gha	ghost	ధ	dha	breathe
ఉ	u	full	ఙ	~ma	sing	న	na	not
ఊ	ua	fool	చ	ca	chain	ప	pa	pot
ఋ	R	Betrri	ఛ	c'a	catch him	ఫ	pha	loophole
ఎ	e	let	జ	ja	jet	బ	ba	ball
ఏ	ae	late	ఝ	jha	hedgehog	భ	bha	abhor
ఐ	ai	lie	ఞ	~na	French n	మ	ma	mother
ఒ	o	rotate	ట	Ta	ten	య	ya	yard
ఓ	oa	rote	ఠ	Tha	ant-hill	ర	ra	run
ఔ	ow	now	డ	d'a	dog	ల	la	luck
అం	am	him	ఢ	dh'a	godhood	వ	va	avert
అ ः	a @h	half	ణ	n'a	under	స	sa	son
			శ	s'a	germanrich	హ	ha	hot
			ష	sha	show	ళ	l'a	Retro L

III. Effect of speaker variant speech:

The performance of an ASR system tested by different speakers' variant tends to markedly degrade as gender and non-native speakers make pronunciation variants as compared to native speakers. In this section the effect of non-native and gender speech on the performance of an ASR system constructed from native speech is discussed. In particular, Telugu is selected the target language for this paper and non-native speaker is from Hindi. To complete this work, a baseline Telugu ASR system is first constructed with CMU phonemes mapped with the Telugu Phonemes are constructed by considering all 51 akshara's of basic Telugu glyphs, and its performance is then evaluated using the Telugu words uttered by the Hindi speakers and Telugu speakers with different region people.

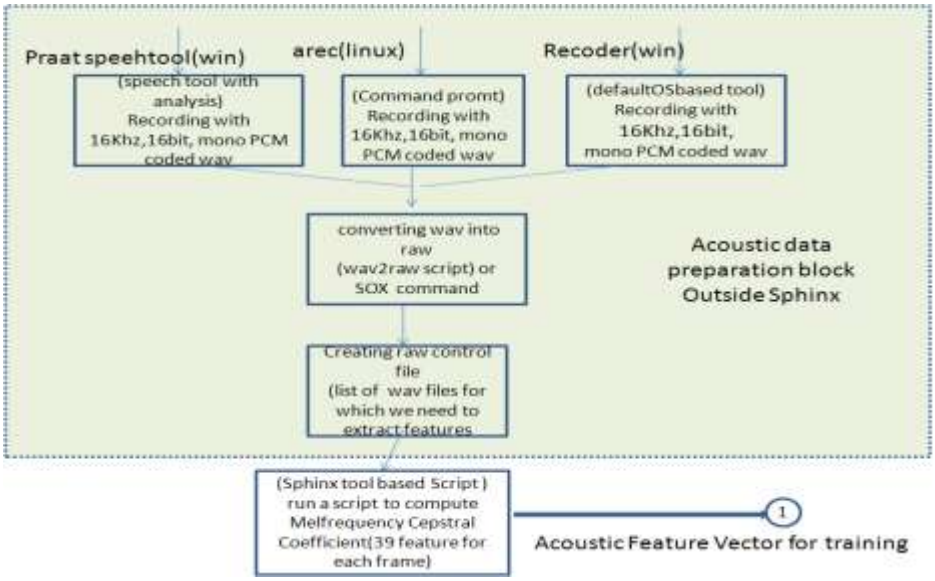
Baseline Telugu ASR

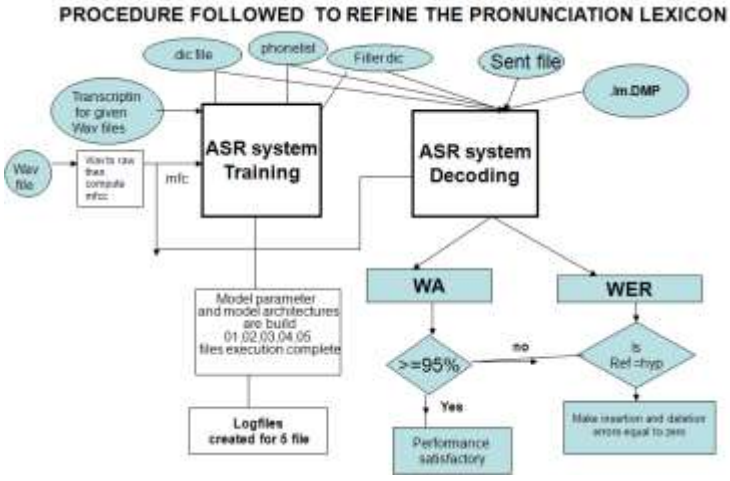
A set of Telugu words are considered(Nagamani, 2010) Telugu wordlist of 665 words is used as the training set for the native-Telugu ASR system. Here after these Telugu word list called UoH word list0. UWL0 is a 665 words of 10 speakers data with regional and nativity variant speech covering syllables of Telugu language word list.

Phonetically rich text corpora

To cover all the symbols or akshara are considered for define the phone list for the Telugu Language. Total 51 akshara are considered hence base form it required 51 symbols of alphabets to represent these phones. First step all the CMU based phones are tried to represent the 51 aksharas. The selected symbols are the derived phone list from the CMU phone list. Human languages are mapped to the Computer language by means of ASCII code. CMU phone list represent 39 phones which are not sufficient and also they are build based on American English pronunciation.

Phonetically rich data can be selected from a medium corpus of text. A set of such words were derived by processing Telugu on-line tutors available in internet and also from primary school syllables of Telugu Language subject to build corpus[3-1]. This process involved translating the Telugu word corpus printed in English(RIT) form to generating set of words that are phonetically rich(i.e., the words contains most phonemes of the language). Here we considered on text based words around 600 which covered most syllables of morphemes and also speaker variability purpose male and female variation study.





Analysis of the pronunciation variability

Pronunciation lexicon for Telugu Speech Recognition system using UOH defined Phonelist. Refine the pronunciation lexicon by adding new pronunciation for the same utterance based on acoustic signal, add or deleted phonemes in the phonelist, refine the acoustic signal or wave file for single speaker.

Vowels:

	VOWELS														
1	AX	AA	IH	IY	UH	UA	RH	EY	IA	AY	OH	OW	AW	AM	AHA
2	AX	AA	IH	IY	UH	UA	RH	IH	IA	AY	OH	OW	AW	AM	AHA
3	AX	AA	IH	IY	UH	UA	RH	EY	IA	AY	OH	OW	AW	AM	AHA
4	AX	AA	IH	IY	UH	UA	RH	IH	IA	AY	OH	OW	AW	AM	AHA
5	AX	AA	IH	IY	UH	UA	RH	AI	IA	AY	O	OW	AW	AM	AHA
6	AX	AA	IH	IY	UH	UA	RH	AI	IA	AY	O	OA	OW	AM	AHA

Consonents:

CONSONENTS																			
K	KH	G	GH	NYA	C	CH	J	JH	INY	T	TTT	D	DDD	NH	TH	TTH	DH	DDH	N
K	KH	G	GHA	NYA	CH	CHH	J	JH	INY	T	TH	D	DH	ND	TH	TTH	DH	DDH	N
K	KH	GA	GHA	NYA	C	CHH	J	JH	INY	T	TTT	D	DXH	NH	TH	TTH	DH	DHH	N
K	KH	GA	GHA	NYA	C	CHH	J	JH	INY	T	TTT	D	DXH	NH	TXH	TTH	DH	DHH	N
P	PH	B	BH	M	Y	R	L	V	SY	S	SH	H	LHA	KSH	RW				
P	PH	B	BH	M	Y	R	L	V	SH	S	SSH	H	LH	KSH	ARA				
P	F	B	BH	M	Y	R	L	V	SH	S	SSH	H	LH	KSH	ARA				
P	F	B	DH	M	Y	R	L	V	SH	S	SSH	H	LH	KSH	ARA				

Table of Confusion pair words in error analysis:

CONFUSION PAIR WORDS AND THEIR PHONEME REPRESENTATION

Reference Word	1	2	3	4	5	6	7	8	9	10	Hypothesis word	1	2	3	4	5	6	7	8	9
JUUDHAGAADUH	J	UA	DH	AX	G	AA	D	UH			DHURAMTHAMUH	DH	UH	R	AX	M	THX	AX	M	UH
DHIYNAARAMUH	DH	IA	N	AA	R	AX	M	UH			GIIRVAANHAMUH	G	IY	R	V	AA	NH	AX	M	UH
PAWLASTYUDUH	P	AW	L	AX	S	THX	Y	UH	D	UH	TAARAAVATHIX	THX	AA	R	AA	V	AX	THX	IH	
BHYBHASTHUDUH	BH	IY	BH	AX	THX	S	UH	D	UH		KIRAATHUDUH	K	IH	R	AA	THX	UH	D	UH	
BUACHIVAADUH	B	UA	C	IH	V	AA	D	UH			THUUDUVILUH	THX	UA	D	UH	V	IH	L	UH	
BHYBHATHSAMUH	BH	IY	BH	AX	THX	S	AX	M	UH											
MUKKARAMUH	M	UH	K	K	AX	R	AX	M	UH		SUUKARAMUH	S	UA	K	AX	R	AX	M	UH	
MAWKTHKAMUH	M	AW	K	THX	IH	K	AX	M	UH		YAAMAVATHIX	Y	AA	M	AX	V	AX	THX	IH	
VAGALAADIX	V	AX	G	AX	L	AA	D	IH												
VAADAKATTUH	V	AA	D	AX	K	AX	T	T	UH		VAALAKAMUH	V	AA	L	AX	K	AX	M	UH	
VIBHAVAMUH	V	IH	BH	AX	V	AX	M	UH			VITHHANAMUH	V	IH	THX	THX	AX	N	AX	M	UH
SHAARDHUULAMUH	SSH	AX	R	DH	UH	L	AX	M	UH		SAANIKOMDAX	S	AA	N	IH	K	O	M	D	AX

IV. Result analysis

The ASR system recognition is performed by taking the Vowel sounds, consonant sounds and Isolated words of different size of letters present in it. Each experiment is carried out by correcting the Substitutional errors. Substitution errors are caused by Insertion and deletion errors will be corrected by adjusting the phone set and acoustic signal and finally reached to 100% word accuracy in individual and combination of Vowel and consonant test data. The confusion matrix for vowel and consonants shown in Figure 4.1 and Figure 4.2 will give the substitution of phones list in lower and upper triangles shown in the matrix. The diagonal will represent the recognition accuracy.

1.1.1 Confusion matrix for Vowels

Figure 4.1. confustion matrix drawn for Vowels in 10 experiments

1.1.1.1 Confustion matrix for Consonents

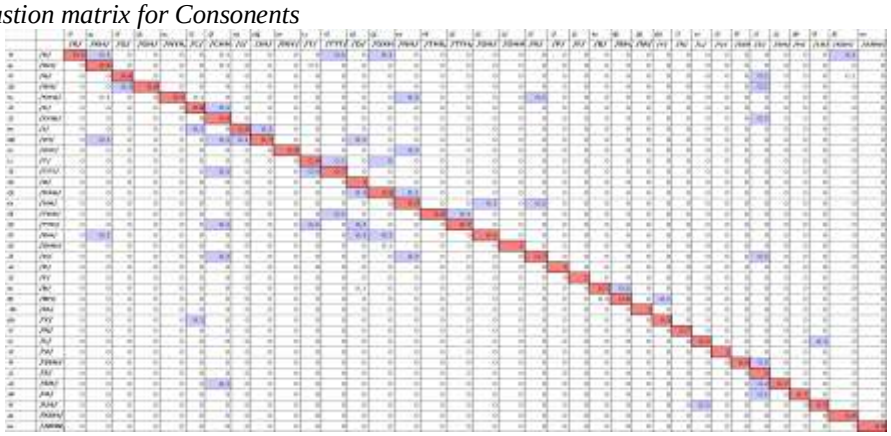


Figure 4.2 Confusion matrix for Telugu Consonents.

References:

[1] "Automatic Speech Understanding" http://ewh.ieee.org/r10/bombay//news6/AutoSpeechRecog/ASR.htm

[2] http://www.research.ibm.com/thinkresearch/pages/2002/20020918_speech.shtml

[2] http://cslu.cse.ogi.edu/asr/

[3] http://www-4.ibm.com/software/speech/

[4] *A. Gallardo-Antolín*, J. Ferreiros, J. Macías-Guarasa, R. de Córdoba and J. M. Pardo* "Incorporating multiple-hmm acoustic modeling in a Modular large vocabulary speech recognition system in Telephone environment" ICSIP2000.

[5] Yoo Rhee Oh, Jae Sam Yoon, Hong Kook Kim "Acoustic model adaptation based on pronunciation variability analysis for non-native speech recognition" *Speech Communication, Volume 49, Issue 1, January 2007, Pages 59-70*